The Journal of Neuroscience, March 1992, 12(3): 895–905

Alternative Splicing of Micro-Exons Creates Multiple Forms of the Insect Cell Adhesion Molecule Fasciclin I

Linda McAllister,[1,a] E. Jay Rehm,[1] Corey S. Goodman,[1] and Kai Zinn[1,2,b]

[1]Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720 and [2]Division of Biology 216-76, California Institute of Technology, Pasadena, California 91125

Fasciclin I is a homophilic cell adhesion molecule in insects that is dynamically expressed on a subset of axon pathways in the embryonic nervous system, and on a variety of other cells and tissues during development. The fasciclin I protein consists of four homologous 150 amino acid domains. In this article, we describe the complete sequence of the *Drosophila fasciclin I* (*fasI*) gene. The gene consists of 15 exons and is distributed over 14 kilobases of DNA. We examine the structure and temporal expression pattern of multiple fasciclin I mRNAs that differ in the lengths of their 3′ untranslated regions. We also show that a highly conserved sequence at the end of the second domain can be altered by the addition of three or six amino acids that are encoded by two alternatively spliced 9 base pair (bp) micro-exons. In grasshopper fasciclin I mRNAs, there are 9 bp and 6 bp insertions at the same position. The first of these insertions is identical in sequence to the first fly micro-exon. The grasshopper insertions are not found together in the same mRNA, so grasshopper fasciclin I species differ by the addition of three or two extra amino acids to the second domain. The alternatively spliced mRNAs are differentially expressed during embryogenesis, and all three of them are present in nerve cord preparations. We suggest that the amino acids inserted by alternative micro-exon splicing may alter the binding specificity of fasciclin I.

Four different membrane-associated glycoproteins expressed on subsets of fasciculating axons in insect embryos, called fasciclin I, fasciclin II, fasciclin III, and neuroglian, have been characterized and their genes cloned (Bastiani et al., 1987; Patel et al., 1987; Harrelson and Goodman, 1988; Snow et al., 1988, 1989; Zinn et al., 1988; Bieber et al., 1989; Grenningloh et al., 1990). Fasciclin II and neuroglian are members of the immunoglobulin superfamily and have domain structures identical to those of the vertebrate cell adhesion molecules N-CAM and L1, respectively. Fasciclin III, although initially not thought to be homologous to any previously identified molecule (Snow et al., 1989), has recently been found to be a diverged member of the immunoglobulin superfamily (Grenningloh et al., 1990). Fasciclin III can function as a homophilic cell adhesion molecule in transfected *Drosophila* Schneider 2 (S2) tissue culture cell lines (Snow et al., 1989). The fourth protein, fasciclin I, has a sequence that is unrelated to any protein in the current databases (Zinn et al., 1988). Analysis of transfected S2 cell lines shows that fasciclin I can also mediate homophilic cell adhesion and cell sorting (Elkins et al., 1990b).

Among cell adhesion molecules, fasciclin I has a unique structure. It is a 72 kDa glycoprotein with an hydrophobic signal sequence but no transmembrane region. The fasciclin I sequence is composed of four domains of approximately 150 amino acids each, identified by virtue of their weak homology (7–15% identity) to each other and by the presence of more highly conserved amino acid "repeats" (up to 45% identity between repeats) at the end of the second, third, and fourth domains, referred to here as D2R, D3R, and D4R. D2R and D3R are about 40 amino acids in length, and D4R is a sequence of about 27 amino acids that corresponds to the N-terminal sections of the other repeats. The repeat sequences have also been highly conserved between grasshopper (*Schistocerca americana*) and *Drosophila*. The fasciclin I protein sequence is only 48% identical overall between the two species, but the repeats are up to 77% identical in sequence, suggesting that they serve an important function, possibly as a binding site for homophilic adhesion (Zinn et al., 1988).

Fasciclin I is attached to the cell surface by a glycosyl-phosphatidylinositol (GPI) linkage (Hortsch and Goodman, 1990). GPI-linked proteins are made as precursors with C-terminal extensions that are cleaved off coincident with attachment of the lipid tail. The cleaved, mature fasciclin I molecule is likely to end 13–15 amino acids C-terminal to the end of the conserved D4R sequence. In *Drosophila* embryos, 40–75% of the protein is also found in a soluble form. This soluble fasciclin I may be produced by endogenous phospholipase activity in the embryo (Hortsch and Goodman, 1990; see below).

Genetic analysis suggests that fasciclin I is involved in the processes of growth cone extension and/or guidance. Flies bearing an apparent null mutation in the *fasciclin I* (*fasI*) gene are viable and show no gross defects in the embryonic CNS or PNS axon arrays. However, embryos doubly mutant in *fasI* and in *abl*, the *Drosophila* homolog of the Abelson tyrosine kinase proto-oncogene, display major defects in CNS axon pathways (Elkins et al., 1990a). These defects are particularly evident in

Received July 22, 1991; revised Oct. 14, 1991; accepted Oct. 18, 1991.

We thank Ann Begovitch and Jayne Danska for help and advice with the PCR experiments, Beth Blankemeir for technical assistance in sequencing the PCR products, Ken Burtis for his *Drosophila* codon bias database, and John Tamkun and Matt Scott for the cosmid library. K.Z. was supported by a Helen Hay Whitney postdoctoral fellowship for part of the period during which these experiments were performed. This work was supported by the Howard Hughes Medical Institute and by grants from the NIH to C.S.G. and K.Z.

[a] Present address: Department of Internal Medicine, University of California at San Francisco, San Francisco, CA 94143.

[b] Present address: Division of Biology 216-76, California Institute of Technology, Pasadena, CA 91125.

the commissural tracts and may be due to a failure of the pioneer growth cones to extend toward the midline of the CNS (Klambt et al., 1991).

In the present study, we carried out a detailed molecular analysis of the gene encoding fasciclin I in *Drosophila*. The *fasI* gene gives rise to transcripts ranging in size from 3 to 5.2 kilobases (kb), which have different temporal and spatial expression patterns. We also demonstrate alternative splicing of 6–9 base pair (bp) micro-exons at identical positions in the *Drosophila* and grasshopper fasciclin I genes. The alternatively spliced mRNAs encode proteins that differ within the most evolutionarily conserved region of fasciclin I.

Materials and Methods

Molecular biology techniques. The cosmid library screen, mapping of clones, and subcloning were performed according to standard methods (Maniatis et al., 1982). The cosmid library was provided by J. Tamkun and M. Scott (University of Colorado). The genomic DNA used for the generation of the library derives from an isogenic *Drosophila melanogaster* stock. The DNA was inserted into the NotBamNot-CoSpeR vector. Twenty thousand colonies were screened using a ^{32}P-labeled fragment of fasciclin I cDNA. Six independent cosmid clones were characterized that span 60 kb, 40 kb of which were extensively mapped; these are called 18-1, 18-2, 18-3, 1-1, 6-1.

Primer extension. Primer extension was performed as described by Bowtell et al. (1988). Two different primers were annealed to RNA from two different stages of development (0–3 and 12–15 hr after egg laying). The primers are indicated in the genomic sequence (see Fig. 2). No difference was seen in the products generated with the two RNA samples.

DNA sequencing. DNA sequencing was carried out according to the dideoxy chain termination method (Sanger et al., 1977) using Sequenase (United States Biochemical Corp.) and ^{35}S-dATP. A 15.3 kb fragment of the genomic region was sequenced, extending from an XhoI site 775 nucleotides (nt) upstream of the start of transcription to an EcoRI site ~1.6 kb downstream from the 3′ end of the previously sequenced cDNA clone. Four fragments were subcloned into Bluescript plasmids (Stratagene Cloning Systems). From 5′ to 3′ they are (1) a 4.3 kb BamHI-XhoI fragment in KS+ ("KBX49"), (2) a 1.9 kb RI fragment in SK+ ("SR1.9"), (3) a 2.6 kb RI fragment in SK+ ("SR2.6"), and (4) a 6.5 kb RI fragment in SK+ ("SR6.5"). Only the first two fragments overlap; the others are contiguous. To confirm the junctions, a 6 kb HindIII fragment that spans the other two junctions was subcloned into SK+ ("SH 6.2") and double-stranded sequencing with specific oligonucleotide primers was used to confirm ~100 nt across each junction on both strands. Three of the four plasmids (1, 2, and 4) were sequenced by a shotgun method in which the supercoiled DNA was randomly sheared by sonication and the ends were repaired with T4 DNA polymerase and Klenow fragment (Maniatis et al., 1982). Fragments in the 400–1000 bp size range were electroeluted from a 1% agarose gel and inserted into the SmaI site of M13mp10 by standard methods. The fourth fragment was sequenced on both strands using specific primers.

RNase protection experiments. RNase protection experiments were performed as described by Zinn et al. (1983). The type III riboprobe was made from a PCR-generated cDNA fragment subcloned into Bluescript. The 3′ riboprobe was generated from the "SR6.5" genomic subclone used for sequencing. For both experiments, 10^6 cpm of probe were hybridized to 30 μg of total *Drosophila* embryonic RNA overnight at 45°C.

RNA preparation and Northern blotting. Total RNA was prepared from staged *Drosophila* embryos according to Crews et al. (1988), separated by electrophoresis on formaldehyde containing 1% agarose gels, and blotted onto nylon membranes. Nerve cord RNA was prepared from bulk isolated embryonic nerve cords prepared according to the "Mash" procedure of Goodman et al. (1984). The RNA was cross-linked to the membrane under UV light, and hybridization to ^{32}P-labeled probes was carried out according to the method of Church and Gilbert (1984).

PCR. Reactions were carried out according to the conditions described by Saiki et al. (1988), using a DNA Thermal Cycler (Perkin Elmer Cetus). Oligonucleotide primers (22–23 nt long) were synthesized based on the cDNA sequence. PCR1a is the sense strand of 298–320 nt in Figure 2, PCR1b is the antisense strand of 1129–1151 nt and PCR2a is the sense strand of the same sequence, and PCR2b is the antisense strand of 1966–1987 nt. The PCR products were digested with restriction enzymes to yield blunt-ended fragments containing the D2R region. The desired fragments were purified from agarose gels, ligated into SmaI-cut M13mp10, and sequenced.

Results

Molecular characterization of the Drosophila fasciclin I *gene*

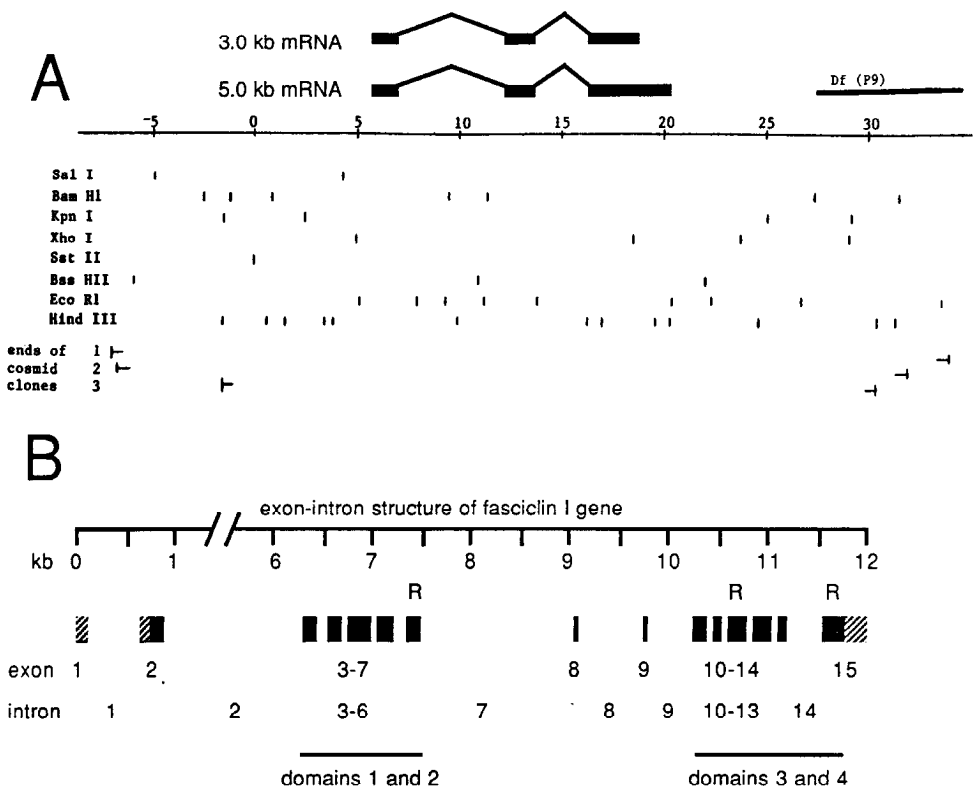
The *Drosophila fasI* gene is located at position 89D on the polytene chromosomes, just proximal to the location of the *Ubx* gene, and is transcribed in the opposite direction from *Ubx* (Elkins et al., 1990a). Hybridization of fasciclin I cDNA probes to restriction digests of cosmids from this region indicated that the 3.0 kb of cDNA sequence is distributed over 12 kb of genomic DNA (Fig. 1*A*). As described below, larger fasciclin I transcripts hybridize to an additional 2 kb of sequence 3′ to this 12 kb segment. A KpnI fragment containing about 3.3 kb of DNA upstream and 5 kb downstream of the transcribed region completely rescues fasciclin I expression in transgenic *fasI* flies (Elkins et al., 1990a), indicating that all the sequences required for correct fasciclin I expression are within this 23 kb DNA segment.

Between the *fasI* and *Ubx* genes, we have found three transcripts in embryonic RNA (data not shown). One of these may correspond to the lethal complementation group mapped to the left of *Ubx* but within Df(P9) (Sanchez-Herrero et al., 1985). The breakpoint of Df(P9) lies about 7 kb from the 3′ end of the *fasI* gene (Bender et al., 1983). In the 12 kb region 5′ to the *fasI* transcribed region there is one embryonic transcript (data not shown), and a portion of the gene corresponding to this transcript may be included within our *fasI* sequence (Fig. 2).

Fifteen kilobases of the *fasI* genomic region were sequenced (Fig. 2). These data, schematically depicted in Figure 1*B*, indicate that the 3.0 kb fasciclin I cDNA sequence is interrupted by 14 introns. Eight of these introns are less than 100 bp in length (ranging from 59 to 81 bp), five are between 0.4 and 1.6 kb, and one intron spans 5.4 kb. The large intron breaks after the codon encoding the 18th amino acid of the mature protein, and is present in the grasshopper gene at the identical position (Snow et al., 1988). Within this intron is the insertion site of the large transposable element, TE77, which creates the *fasI^{TE}* mutation (Elkins et al., 1990a).

One of the larger introns breaks between the sequences encoding the second and third 150 amino acid domains to split the protein coding region into two equal halves. Introns are not located at the other interdomain junctions. The largest exon includes the fourth domain conserved repeat (D4R), the GPI attachment sequence, and the 3′ untranslated region. Eight out of 13 of the introns in the coding region fall after the third nucleotide of the codon.

The start of transcription of the gene was determined by primer extension. These results indicate that there is a single start site 17 nt upstream of the 5′ end of the sequenced 3.0 kb cDNA (data not shown). This is the same start site seen in *in vitro* transcription experiments using a restriction fragment spanning the start site (M. Biggin and L. McAllister, unpublished observations). The sequence surrounding the start of transcription does not closely resemble the *Drosophila* consensus for start sites (ATCA G/T T C/T; Hultmark et al., 1986). The best match for the *Drosophila* TATA box consensus (GCTTTAAAAGCC; Bray and Hirsh, 1986) is GCTTAAACC and occurs between positions −18 and −27. Various repeated sequences, including



C

SPLICE DONOR			SPLICE ACCEPTOR	
CTG GTACTT	IVS1	626NT	TGCCTTTGAATTGCAG	A
CAG GTGAGT	IVS2*	5692NT	CCGTTTATCTTTGCAG	T
CAG GTGAGT	IVS3	81NT	TGATTACTCTTTGCAG	C
CAG GTCAGT	IVS4*	74NT	CATTACTTACCGACAG	A
AAG GTGCTT	IVS5*	67NT	TGCTTTATGACTCCAG	C
AAA GTAAGA	IVS6	77NT	ATAATTTTACTTGCAG	T
CAG GTATGT	IVS7*^	1562NT	AAACCAAAACGTTCAG	T
AAG GTAAAC	IVS8*^	793NT	TTTAAAACTGATTTAG	T
AAC GTAAGT	IVS9*^	454NT	TATCTTGAAGGAGAAG	G
GCG GTAATG	IVS10	66NT	TAAATACATCTTACAG	A
TTG GTGAGC	IVS11*	63NT	CCATTCTTTTCAAAAG	A
GAG GTATGT	IVS12	59NT	TATCGTATTCCTTAAG	T
CAT GTAGGT	IVS13*	60NT	TGTTTTGTAACCGCAG	T
GAC GTAAGT	IVS14	409NT	TACTTCCATTTTGCAG	G
C AG GTRAGT A	CONSENSUS SEQ		YYYYYYYYYYYYNCAG	G

Figure 1. Molecular map of the fasciclin I gene. *A*, A map of 40 kb of genomic DNA covered by cosmid clones that hybridize to a cDNA probe. A unique SstII site corresponds to zero on the map. The recognition sites of eight different restriction enzymes are indicated below the scale line. The structures of the two major fasciclin I mRNAs are indicated above the scale line, which is divided in 5 kb segments. Only the two largest introns are shown. The direction of transcription of these mRNAs is from left to right. The DNA segment that rescues normal fasciclin I expression in P-element transformants is bounded by the KpnI sites at +2.5 kb and +25.5 kb (Elkins et al., 1990a). The sequenced genomic DNA extends from the XhoI site at +5.0 kb to the EcoRI site at +20.3 kb. The sequences removed by the Df(P9) deletion are also shown. The borders of three of the cosmid clones (18-1, 18-2, 18-3, labeled *1, 2, 3,* respectively) are indicated at the bottom. *B*, The 3.0 kb fasciclin I cDNA sequence spans 13 kb within the 15.3 kb of sequenced genomic DNA. There are 15 exons, indicated by *solid* and *hatched bars*; the *solid segments* designate coding regions. The exons containing the highly conserved repeat sequences are indicated above by *R*. On the line below are the exon numbers, and below this the intron numbers. The regions containing exons encoding domains 1+2, and domains 3+4, are also shown. *C*, The splice junctions for the 14 introns (IVSs) are compared to the *Drosophila* splice site consensus (Burtis and Baker, 1989). The regulated junctions (marked "^") have weak acceptor sites with fewer pyrimidines (4–6/12 vs. 7–10/12) preceding the conserved CAG splice site. The 8/13 coding region IVSs (*) occur after a codon.

a candidate homeobox protein binding site, are found in the 5′ flanking sequences and are indicated in Figure 2.

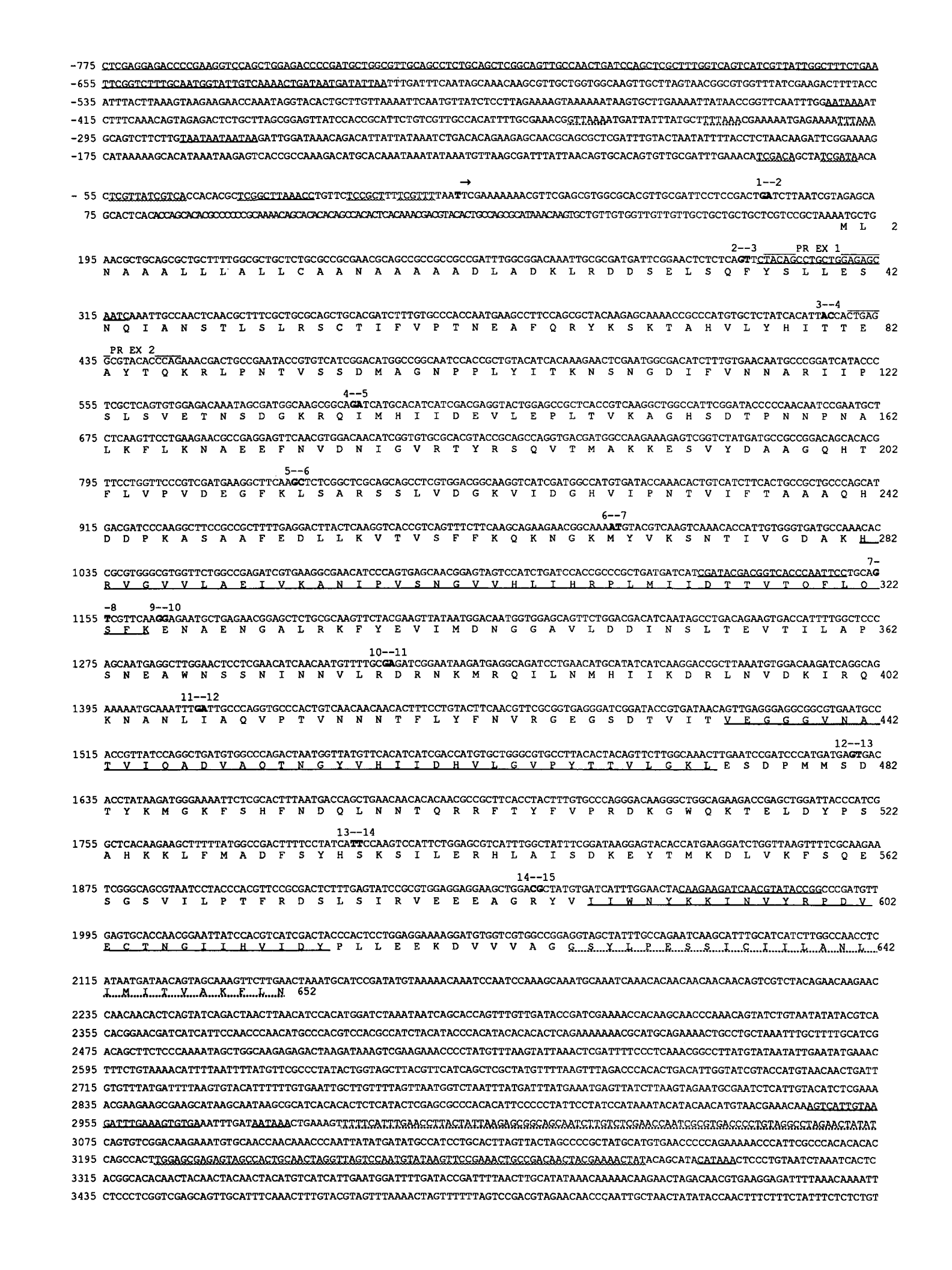
The known splice junctions in the *fasI* gene are compared in Figure 1*C* to the *Drosophila* splice site consensus (Burtis and Baker, 1989). The splice donors are uniformly similar to the consensus. However, the splice acceptors of the alternatively spliced junctions described below are less similar to the consensus than are the other acceptors in the *fasI* gene. Specifically, IVS 7-9 have fewer pyrimidines preceding the splice site than do the rest of the acceptors (4-6/12 vs. 7-10/12). Such "weak acceptor sites" occur at other known regulated splice junctions (Brown et al., 1989; Burtis and Baker, 1989).

Expression of fasciclin I mRNAs during development

Both the grasshopper and *Drosophila* fasciclin I cDNAs, when hybridized to blots of RNA from the appropriate species at similar stages of embryogenesis, recognize a band of 3.0–3.2 kb, approximately the length of the sequenced cDNAs (Fig. 3*A*). In grasshopper, a 5.2 kb transcript is also observed. In *Drosophila*, there is a second major transcript of 5.0 kb, as well as three minor transcripts of 4.0, 4.2, and 5.2 kb.

The temporal pattern of expression of the fasciclin I mRNAs during *Drosophila* embryogenesis is shown in Figure 3*B*. The 3.0 kb transcript is expressed very early, suggesting that it could be a maternal transcript. This mRNA is also present in dissected ovaries (Fig. 3*D*). The early expression of fasciclin I mRNA suggests that, as in grasshopper (Bastiani et al., 1987), the protein exists outside the nervous system, and this has been confirmed by immunohistochemistry (data not shown).

During embryogenesis, the 5.0 and 3.0 kb transcripts are expressed at highest levels between 9 and 15 hr of development. This correlates with the time of axon outgrowth during nervous system development (Fig. 3*B*). The minor transcripts, of 4.0, 4.2, and 5.2 kb, can also be detected in RNA from this time



```
3675 TAACTATCAGCCTACCAGTACACTGTCTCAACTCTCACCACCACCACCCACTTAGGAACTCAGTCGAACTTGAACTCGAAATCAAAGATCCAGTTGTGGCAGTCGCTTCACGTAGTT
3795 GCTAATTCCCAATTCGAACCGATCCTTTCCGAAAGTCTTATCTTTAGTATAGGTGGTTTAGTTTCATTTGGAGCCGTGCAGTGCCGTAGCAGCTAAGTAAAAATGTATGAAATGAAGATG
3915 AACACGAGATCGAAATCGTACGGAATGATCAGAAATCAGAATAAATAATGAATACGCTAATGAATTGTACAAGTAAGCTTTAAAGAATTGCTGGAGGAGCGCGGATCGGAGAACTTAGA
4035 GGAGGGAGAACCGCATTGCAATCGCATTGCAATTTGTGTCGTAGTCAGTAGTTACACGTTAAGCGGCGTCTTAACGTGTAACTAGTGCCTTACTAAAGATAAACGCATTACCTTAACCTT
4155 TATACAAATTTACTCAAAACATACTTGTACCCCAAGCATACGTTCCGCTTCGAATGATACCCAGATATATATACGGAGTTACACCCCAAGAATACAAGTATAACTACAAATGATATTGCG
4275 CCACACGCTATTTACACCAAATACACCAAACAAATCGAGAAATGCATATTTTTCATATATTTAATTGTCAGAATAATATAACGTATATGTAATATGTAGTTTATTTACTGTAAAACGCAA
4395 GAACCTAACAAGTGGAATTTGAATCACATACAATTGATGTATATTAGCTATTGAGTTTCTAAGCAAGCGTTAGACACTGAAATATATGTTCAAATATATGATTCAAATATATACAATATG
4515 CGAAACCAAGCAAACTATGGAAAACTGGAGTGCCAAAAGAATATTATTCCATTTTATTTTACGACAAGCGCTTTTTACAAATAAACCGAATCCATTTAAATTACTCGTAAAGAACAGACA
4635 GATTATATTTAGCATTAGTTAAACTAATTATTACATGTACTAGAAAACCGAATGTCAACCGAGAATCTTCAGCAAGCTTGAGCGAATAATAAAACTTTAAAACTAACTATAAATAAATCG
4755 ACCGTCCTTTTTTGCACTTAATCATGGGTTATGGTTGTGGAACCTGAATTC
```

Figure 2. Fasciclin I genomic sequence. The entire 3.0 kb cDNA sequence (Zinn et al., 1988) is shown with the 5′ and 3′ genomic sequence. Nucleotide positions relative to the start of transcription are at the *left*. Although the intron sequences are not shown, their positions are indicated by *double dashes above the sequence,* and the two residues immediately flanking the introns are in *boldface.* The exon numbers are also indicated. The complete sequence has been deposited in the EMBL/GenBank Data Libraries under accession number M32311. The amino acid sequence is indicated *under* the nucleotide sequence, and amino acid numbers are to the *right. Underlined regions,* from the beginning of the sequence, are (1) an open reading frame 5′ to the *fasciclin I* gene; (2) an AATAAA sequence that may be for an mRNA containing this open reading frame; (3) three repeats of the sequence (G/T)TTAAA (*dotted lines*); (4) four contiguous repeats of TAA; similar sequences have been shown to bind homeodomain proteins (for review, see Biggin and Tjian, 1989); (5) seven repeats of a sequence with a consensus of TCGNY(A/T) (*wavy lines*); (6) the *Drosophila* TATA box homology (Bray and Hirsh, 1986); (7) the D2R, D3R, and D4R amino acid sequences; (8) the amino acid region that is probably removed by cleavage coincident with GPI attachment (*dotted line*; the N terminus of this segment is undetermined and could be at another residue near G623); (9) a sequence conserved between the 3′ untranslated regions of the grasshopper and fly cDNAs; (10) the 3.0 kb mRNA polyadenylation signal; (11) probe 3′-1 (*dotted line*); (12) probe 3′-2 (*wavy line*); (13) a potential polyadenylation signal whose use would generate a 3.2 kb mRNA; (14) a polyadenylation signal that could correspond to the 4.0 kb mRNA; (15) three repeats of the sequence (in RNA) AUUUA, which can contribute to message instability (Shaw and Kamen, 1986) and may also participate in translational control (Kruys et al., 1989); (16) probe 3′-3; (17) two polyadenylation signals that would correspond to mRNAs of 4.75 kb; these may be the signals for the "5.0" kb RNA. An *arrow* shows the start of transcription of the *fasI* gene as determined by primer extension and *in vitro* transcription. The actual start is likely to be within 1 bp of the *T* indicated in *boldface.* The primers used in primer extension experiments are indicated above the sequence by *lines* and are labeled *PR EX 1* and *PR EX 2.*

period (Fig. 3*A*). Later, only the 3.0 and 5.0 kb transcripts are seen, and their levels decrease gradually until the end of embryogenesis (24 hr). The 3.0 and 5.0 kb transcripts, as well as an additional 3.2 kb mRNA, not seen in Northern blots from whole embryos, are detected in RNA blots from mass-isolated 10–13 hr ventral nerve cords (Fig. 3*C*).

In first and second instar larvae, the levels of expression of fasciclin I mRNAs are very low. Longer exposures of blots such as the one in Figure 3*C* reveal that there is some expression of the 3.0 kb transcript during these stages. By the third instar, the 5.0 kb RNA is detectable and the 3.0 kb is more abundant (Fig. 3*C*). RNA from mixed male and female adults contains both the 3.0 and 5.0 kb transcripts, but the 3.0 kb RNA is apparently female specific (Fig. 3*D*). This transcript is the only one seen in dissected ovaries. An RNA of the same size as the nerve cord–specific form is also found in adults (Fig. 3*D*). The levels of fasciclin I mRNAs in adult flies are somewhat lower than those found in 9–15 hr embryos; the blot of Figure 3*C,* which shows a darker band in adult RNA, was exposed much longer than that of Figure 3*B.*

In total, we have identified six fasciclin I mRNAs: two major species of 3.0 and 5.0 kb, and four minor species of 3.2, 4.0, 4.2, and 5.2 kb. All of these are expressed in Schneider 1 (S1) cells at high levels (Fig. 3*C*; Elkins et al., 1990a). All six of these RNAs are found in the polyA$^+$ RNA fraction.

Fasciclin I transcripts differ primarily in their 3′ untranslated regions

In order to define the additional sequences present in the longer RNAs, we generated probes for Northern blot hybridization from the 5′ and 3′ regions immediately flanking the 3.0 kb RNA transcription unit, and from within the largest intron, IVS 2. The 5′ and IVS 2 probes do not hybridize to polyA$^+$ RNA from 9–12 hr embryos (data not shown). However, the 3′ probe hybridizes to all of the transcripts except for the 3.0 kb mRNA (data not shown). To map the larger transcripts, three smaller probes from the 3′ region (3′-1, 3′-2, and 3′-3 underlined in Fig. 2) were hybridized to 9–12 hr polyA$^+$ RNA. The two proximal probes, 3′-1 and 3′-2, hybridize to most or all transcripts that are longer than the 3.0 kb mRNA. The most distal probe, 3′-3, is 1.6 kb from the end of the 3.0 kb transcription unit and hybridizes exclusively to the 5.0 kb mRNA. (Fig. 3*E*). Probes located farther away than 3′-3 from the fasciclin I cDNA sequences hybridize to the 3′ flanking transcripts discussed above, which do not contain fasciclin I sequences.

These results indicate that the longer transcripts extend farther in the 3′ direction, and that this 3′ extension can account for the size differences. We were unable to isolate cDNAs containing sequences from this region, even after screening more than a million clones from several libraries with 3′ probes specific for the larger RNAs (the abundance of the 3.0 kb cDNA is about 1 in 10,000 clones). Thus, we do not know if there is additional splicing within this region. The sequence of the 3′ untranslated region is devoid of open reading frames that display *Drosophila* codon usage. There are four additional polyadenylation signals downstream from the end of the 3.0 kb cDNA. In conclusion, the major length differences between the multiple mRNAs can be accounted for by the use of different polyadenylation sites giving rise to additional 3′ untranslated sequences.

In S1 cells, all of the fasciclin I protein is attached to the membrane by a GPI linkage. In contrast, much of the protein found in *Drosophila* embryos is in a soluble form, and this percentage changes during development (Hortsch and Goodman, 1990). We were interested in determining whether some of this soluble fasciclin I is translated from a variant mRNA species that lacks the GPI attachment sequence. To determine whether such an RNA exists, we looked directly at the last exon of the gene (exon 15), which includes the GPI linkage signal. An RNase protection experiment was performed with a riboprobe covering the genomic region encoding the carboxy ter-

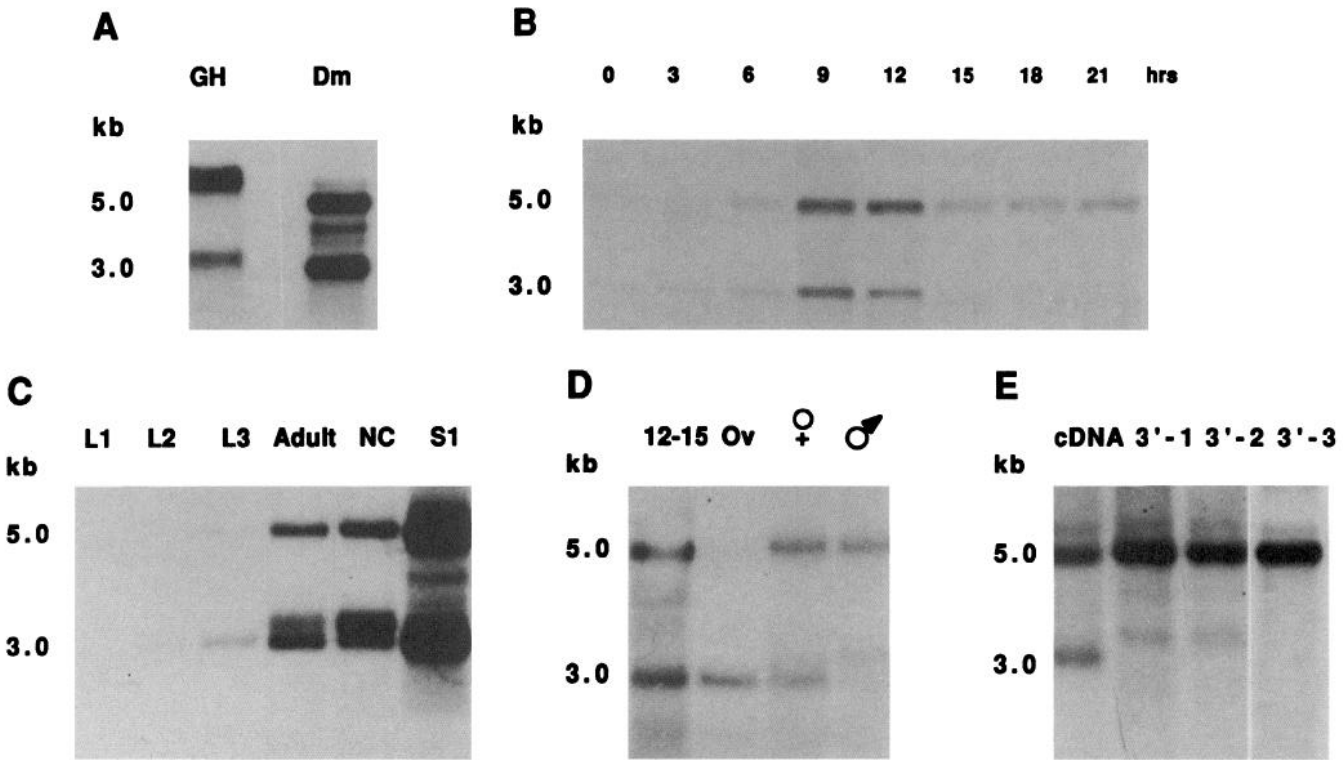


Figure 3. Northern blot analysis of fasciclin I transcripts. *A*, PolyA$^+$ RNA (10 μg) from 35–50% grasshopper embryos (*lane GH*) or total RNA (15 μg) from 9–12 hr old *Drosophila* embryos (*lane Dm*) was subjected to gel electrophoresis, transferred to a nylon membrane, and probed with coding region fragments from the corresponding cDNAs. The positions at which 5.0 and 3.0 kb RNAs migrate are indicated to the left of each panel. *B*, Total RNA (15 μg) from the indicated period in embryogenesis was analyzed as in *A*. Each number corresponds to a 3 hr period beginning at that number of hours (i.e., *lane 9* contains RNA from 9–12 hr old embryos). *C*, Total RNA (15 μg) from the indicated larval instar (*lanes L1–L3*), from adult flies, from mass-isolated 9–12 hr embryonic ventral nerve cords (*lane NC*), and from Schneider 1 tissue culture cells (*lane S1*) was analyzed as in *A*. *D*, Total RNA (15 μg) from 12–15 hr old embryos, from dissected ovaries (*lane Ov*), and from female and male adult flies (*lanes labeled with symbols*) was analyzed as in *A*. *E*, Total RNA (15 μg) from 9–12 hr old embryos was analyzed with the indicated probes, which are described in Materials and Methods and in the Figure 2 caption.

minus of the protein plus 700 nt of 3′ untranslated region (exons 10–15, Fig. 4*A*). If alternate splicing occurred within the coding region of exon 15, it would be expected to generate a fragment of 200 nt or less. No such fragment is observed with either 0–3 or 12–15 hr RNA (Fig. 4*B*). This experiment also shows that there is no detectable alternative splicing of exons 10–14. Thus, it is unlikely that the soluble form of fasciclin I is encoded by a separate mRNA.

PCR reveals alternative splicing of micro-exons in grasshopper and Drosophila

A striking feature revealed by comparing the genomic sequence to the original cDNA sequence (Zinn et al., 1988) is the presence of nine extra nucleotides in the cDNA at the end of the second domain conserved repeat (D2R) that are not found in the exons flanking this point. These missing nucleotides (TCGTTCAAG, encoding the amino acids SFK) are found within intron 7 and are flanked by AG and GT sequences, indicating that they comprise a separate micro-exon.

To investigate the splicing of this micro-exon, we used PCR analysis (Saiki et al., 1988) to examine the structures of the fasciclin I mRNAs within this region. Primers were used that lie within D2R (exon 7) and D4R (exon 15). When these primers are used in PCR reactions with the 3.0 kb cDNA as template, an 850 bp product is observed. A single product of this size is also observed after PCR amplification of embryonic cDNA from several time points. A single band was also obtained when primers in exons 3 and 7 were used for amplification. These data indicate that no large exons are skipped or added in fasciclin I mRNAs. However, when the amplified exon 7–exon 15 fragment was cloned and individual isolates sequenced, three different products were found. Two of these differ from the cDNA clone in the region of D2R. The shortest fragment does not include the nine nucleotide micro-exon (exon 8), and consequently exons 7 and 10 are contiguous. The second fragment corresponds to the cDNA sequence, in which exons 8 and 10 are contiguous. The largest fragment includes exon 8 followed by an additional 9 nt sequence (TTTATGAAC), which is in turn followed by exon 10. Examination of the genomic sequence shows that these additional 9 bp arise from a second micro-exon (exon 9), found in the intron between exons 8 and 10. This micro-exon encodes the amino acids FMN. Thus, three fasciclin I mRNAs are generated by the alternative splicing of two micro-exons of 9 nt each. The cDNA lacking either micro-exon is called the type I cDNA, the type II cDNA includes the first micro-exon (exon 8), and type III includes both exons 8 and 9 (Fig. 5). Eighty-eight clones were sequenced from PCR-amplified 9–12 hr cDNA, and 72 of these were type I, 10 were type II, and 6 were type III.

The D2R amino acid repeat is 77% identical between grasshopper and *Drosophila* (Zinn et al., 1988), suggesting that it serves an important function. We were interested to see if the

grasshopper fasciclin I mRNAs also showed microheterogeneity at the end of D2R. PCR was used to generate a product including this region of the grasshopper fasciclin I sequence from 35–50% embryo RNA. Three types of grasshopper PCR products were identified. One corresponds to the *Drosophila* type I cDNA (no micro-exons), and another includes a sequence at the end of D2R encoding the identical micro-exon encoded amino acids (SFK) found in the type II *Drosophila* cDNA. The third grasshopper form, which we call type IV, has a six nucleotide insertion (GGTTTT) at the end of D2R that encodes the amino acids GF (Fig. 5). This type IV cDNA corresponds to the grasshopper fasciclin I clone that was originally sequenced (Zinn et al., 1988). Fifty-one grasshopper PCR clones were sequenced, and 42 type I cDNAs, 5 type II cDNAs, and 4 type IV cDNAs were found. No clones were found that would correspond to the type III cDNA of *Drosophila,* which contains both micro-exons. Since the second insertion in the grasshopper fasciclin I mRNAs is not identical in sequence to the second fly micro-exon, we examined the *Drosophila* genomic sequence within introns 7, 8, and 9 for a possible exon encoding GF. None was found.

Thus, in *Drosophila,* alternative splicing at the D2R junction can extend the conserved repeat by three (type II) or six amino acids (type III). In grasshopper, the repeat can be extended by three (type II) or two (type IV) amino acids. We include the alternatively spliced sequences in the repeat since the amino acids encoded by exon 8 (SFK) are the same in both grasshopper and *Drosophila* (Fig. 5). Heterogeneity is unlikely to occur by this mechanism in the other repeats (D3R and D4R) since they are not immediately followed by introns (Fig. 2). This has been confirmed in *Drosophila* by PCR for D3R (this region was also contained within the PCR products) and by RNase protection across the exon encoding D4R (Fig. 4).

Alternative splicing of micro-exons is regulated during development

The developmental pattern of expression of the two 9 bp micro-exons was determined by RNase protection. An RNA probe made from the D2R junction region of a type III cDNA fragment (containing both micro-exons; see Fig. 6*A*) can detect all three types of fasciclin I mRNA. Analysis of RNA from various stages shows that the type I mRNA predominates throughout embryogenesis (Fig. 6*B*). The type III message is also expressed throughout embryogenesis but at very low levels. It is most prevalent at 9–15 hr. The type II mRNA is expressed almost exclusively at 9–15 hr. All three types are expressed in RNA from purified 10–13 hr ventral nerve cords. This experiment also allows us to look for a potential transcript containing only exon 9, the second micro-exon, at the D2R junction. No product corresponding to such a transcript was seen, even after long exposure times.

A similar experiment was performed with RNA prepared from larvae, dissected ovaries, adult females, and adult males (data not shown). No type II mRNA was detected in these samples. As in early and late embryogenesis, the type I mRNA predominates and type III is expressed at very low levels.

Comparison of these results with the Northern blot analysis (Fig. 3) indicates that there is unlikely to be a strong correlation between polyadenylation site selection and micro-exon splicing. The 3.0 and 5.0 kb *Drosophila* fasciclin I mRNAs are present at approximately equal levels at 9–15 hr, but the abundance of type II and III messages is much less than that of type I. This suggests that both the 3.0 kb and 5.0 kb RNAs can have the

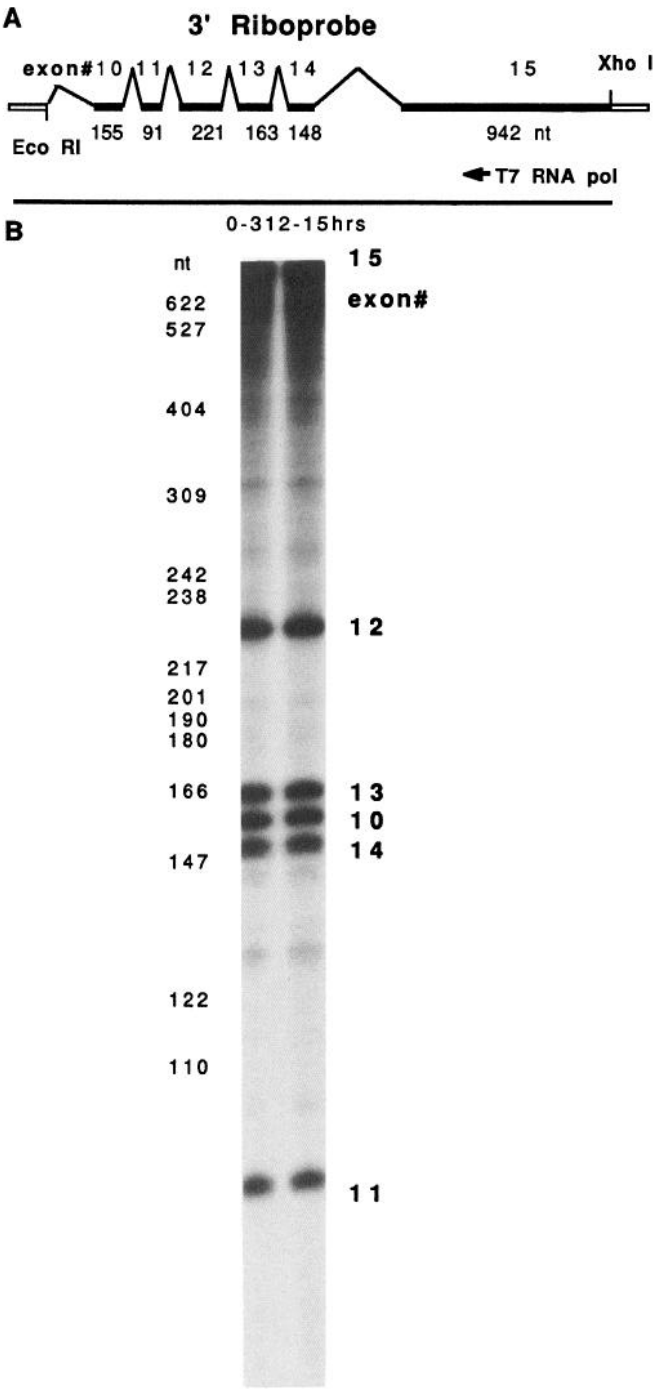


Figure 4. RNase protein analysis of fasciclin I 3′ exon splicing. *A,* Schematic diagram of the fasciclin I gene splicing pattern and of the sequences contained in the riboprobe. The probe was generated by transcription with T7 RNA polymerase and contains sequences complementary to exons 10–14 in their entirety, 942 nt of exon 15, introns 10–14, and most of intron 9. *B,* The radiolabeled probe was hybridized to total RNA (30 μg) from 0–3 hr or 12–15 hr old embryos, treated with RNase, and subjected to electrophoresis on a denaturing acrylamide gel (Zinn et al., 1983). Positions of marker DNA fragments of the indicated sizes are on the left. The bands corresponding to particular exon numbers are labeled on the right. Signals for all six exons within the probe are evident, and no other major bands are present.

```
Grasshopper HATGAVLAEIVKANIPVKNGVVHLIQRPLMVVDNTVKQFLE  {SFK}  or  {GF}
            :  : ::::::::::::: ::::::: ::::  : :: :::    :::
Drosophila  HRVGVVLAEIVKANIPVSNGVVHLIHRPLMIIDTTVTQFLQ  {SFK} {FMN}
                                                        II   III  IV
```

Figure 5. Alteration of second domain repeat sequences in grasshopper and *Drosophila* by alternative micro-exon splicing. The amino acid sequences of D2R from grasshopper and fly are indicated. *Double dots between the lines* indicate sequence identities. At the right are indicated, in *braces,* the sequences added to the repeat by the micro-exons in the two species. The *roman numerals* indicate the mRNA type corresponding to the inclusion of the sequences in braces.

type I splicing pattern. In addition, we know from the cDNA sequences (Zinn et al., 1988) that the 3.0 kb transcript can have the type II splicing pattern (in *Drosophila*) or the type IV splicing pattern (in grasshopper).

Discussion

Structure and transcription of the Drosophila fasciclin I *gene*

The 72 kDa fasciclin I glycoprotein was originally identifed in a monoclonal antibody screen for surface proteins expressed on subsets of axons in developing grasshopper embryos (Bastiani et al., 1987). cDNA clones encoding the protein were isolated and used to screen an embryonic *Drosophila* cDNA library (Snow et al., 1988; Zinn et al., 1988). Sequence analysis of these clones showed that fasciclin I is composed of four homologous 150 amino acid domains. The C-terminal portions of the second, third, and fourth domains are highly conserved between species, and we have called these D2R, D3R, and D4R. These sequences are also the regions most highly conserved between domains in a single species (Zinn et al., 1988). Fasciclin I has no transmembrane domain and is attached to the cell surface by a GPI linkage (Hortsch and Goodman, 1990). When expressed on the surface of transfected S2 cells, fasciclin I can mediate homophilic adhesion and cell sorting (Elkins et al., 1990b). Embryos homozygous for a fasciclin I null mutation, *fasI*^{TE}, exhibit no gross defects in the CNS or PNS. However, when *fasI*^{TE} is combined with a mutation in the *Drosophila* homolog of the *abl* tyrosine kinase, a phenotype is produced in which commissural axon pathways in the embryonic CNS are reduced or absent (Elkins et al., 1990a).

In this article, we describe the complete sequence of the *Drosophila fasciclin I* gene, which spans 14 kb of DNA and contains 15 exons. One intron breaks immediately after the second domain, splitting the coding region into two equal halves. Each half (domains 1+2, and domains 3+4) is encoded by six exons. This suggests that a gene encoding an ancestral two-domain protein may have duplicated to produce the gene encoding the four-domain fasciclin I protein.

The gene is transcribed at all stages of development. There are six distinct size classes of fasciclin I mRNAs, which are expressed in a complex temporal and spatial pattern. The 3.0 kb mRNA corresponds to the cDNA clone that was originally sequenced (Zinn et al., 1988). This mRNA is expressed throughout embryonic development. It may have a maternal component, since it is present very early and is also expressed in dissected ovaries. In the adult, the 3.0 kb mRNA is female specific and may be confined to the ovaries. The other major species, a 5.0 kb mRNA, probably does not have a maternal component. Both the 3.0 and 5.0 kb mRNAs are expressed at maximal levels between 9 and 15 hr of embryogenesis. This is the period during which axon outgrowth occurs. The 5.0 kb mRNA differs from the 3.0 kb mRNA only in the length of its 3′ untranslated region and therefore encodes the same fasciclin I protein.

In addition to the major mRNAs, there are also four minor species, of 3.2, 4.0, 4.2, and 5.2 kb. These also differ from the 3.0 kb species only in the length of their 3′ untranslated region. The 3.2 kb mRNA is enriched in mass preparations of embryonic ventral nerve cord. All of the mRNAs are probably generated through the use of different polyadenylation sites. There are several potential polyadenylation signals in the 2 kb of sequence 3′ to the end of the 3.0 kb cDNA that could correspond with these mRNAs. All of these mRNA size classes apparently contain the same coding region(s), so it is unclear why the transcription pattern of the gene is so complex. *In situ* hybridization to whole-mount embryos with probes from the coding region and from the 3′ region specific to the 5.0 and 5.2 kb RNAs shows that the spatial expression patterns of the shorter and longer RNAs are similar or identical (L. McAllister, C. S. Goodman, and K. Zinn, unpublished observations). It is possible that the different size classes have different half-lives, and this may be important in controlling the amount of fasciclin I protein present on an individual cell at a particular time in development.

The expression pattern of fasciclin I is very dynamic, and the protein is rapidly eliminated from specific tissues at several stages during development (K. Zinn, L. McAllister, and C.S. Goodman, unpublished observations). This could be due to a combination of short mRNA half-life and cleavage of fasciclin I from the membrane by a phospholipase. Hortsch and Goodman (1990) have found that most of the fasciclin I in the embryo is in the soluble fraction at certain times during development. The data presented here indicate that this soluble fasciclin I is not produced from a different mRNA that lacks the GPI attachment sequence, and therefore must be the result of phospholipase cleavage. The existence of a soluble form of fasciclin I is consistent with ultrastructural observations of fasciclin I immunoreactivity in the spaces around developing axons and glia (Bastiani et al., 1987). Local phospholipase cleavage of fasciclin I could be important for defasciculation at choice points. It is also possible that soluble fasciclin I could bind to the extracellular matrix and thus play a role in interactions between growth cones and the matrix substrate.

Fasciclin I protein diversity is generated by alternative splicing of micro-exons

Although the distinct size classes of *Drosophila* fasciclin I mRNAs all seem to produce the same protein(s), we have found microheterogeneity in the mRNA sequences encoding the junction between the second and third domains of fasciclin I. One of the larger introns breaks at the end of D2R. The 3.0 kb cDNA clone that we originally sequenced contains three amino acids at this position that are not encoded by the large exons (7 and 10) that flank the intron. These three amino acids, SFK, are encoded by a separate micro-exon (exon 8) of 9 bp. In order to determine whether this micro-exon is always included in fasciclin I mRNAs, we performed PCR analysis of embryonic cDNA. These data

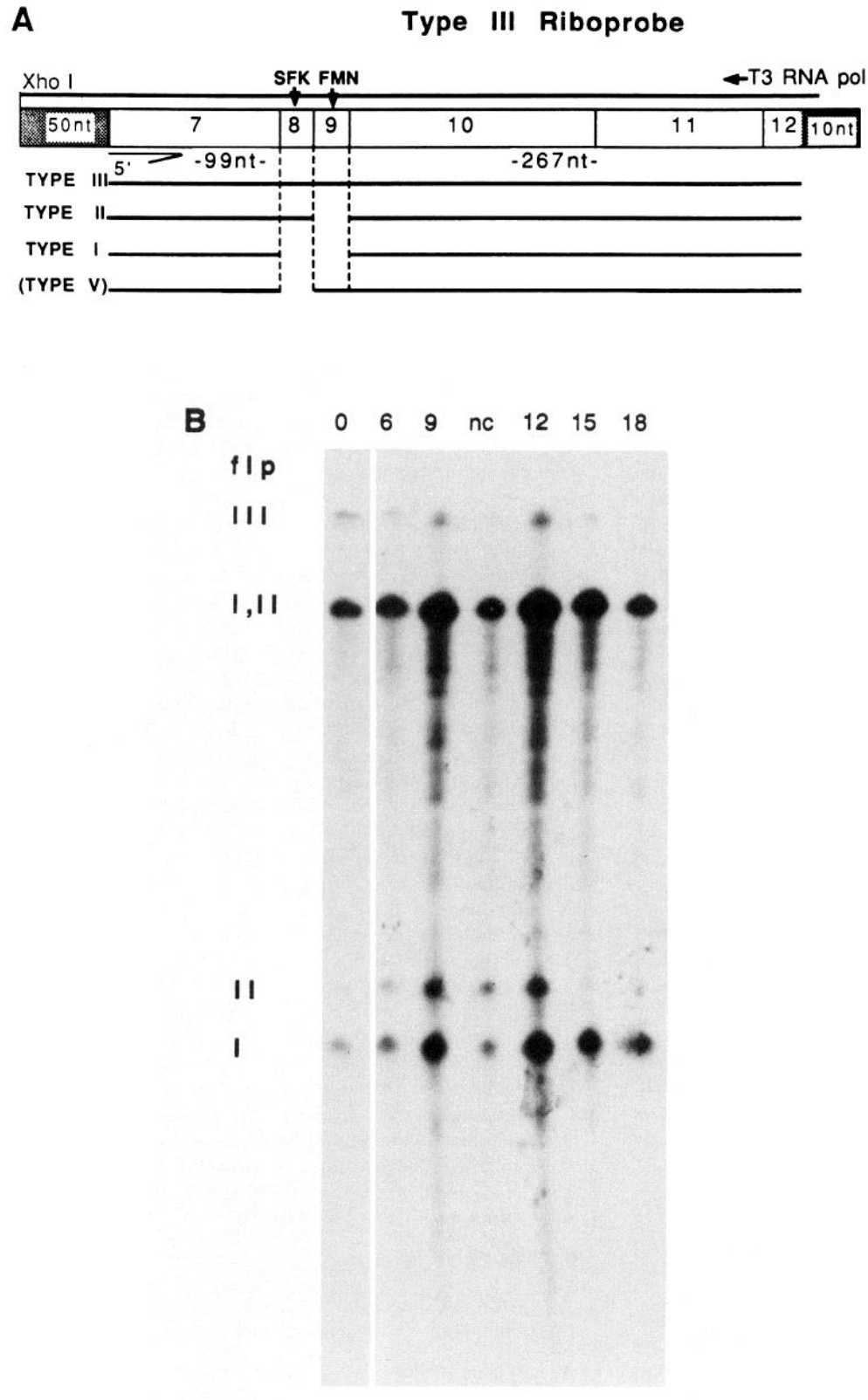


Figure 6. RNase protection analysis of alternative micro-exon splicing during development. *A,* A type III cDNA fragment, that is, one containing exons 8 and 9, was generated by PCR and subcloned into the transcription vector. The exons of the fasciclin I gene are indicated in the *boxes,* as are vector sequences, which are *shaded.* The amino acid sequences of the two micro-exons are shown above the corresponding boxes. The direction of *fasciclin I* gene transcription is from left to right. The full length probe is 444 nt long and includes parts of exons 7 and 12, all of exons 8–11, plus an additional 59 nt of vector sequence. The products that would be generated by RNase protection with this probe of fasciclin I mRNAs of the indicated types are schematically depicted by *lines.* Except for the type III mRNA, each species will generate two bands. The products that would be produced by the hypothetical type V RNA (micro-exon 9 only) are also indicated. The lengths of the products are indicated *above* the lines. *B,* Radiolabeled probe was synthesized by transcription with T3 RNA polymerase, hybridized to total RNA (30 μg) from the indicated stages, and analyzed as described in the Figure 4 caption. *Lane 0,* RNA from 0–3 hr old embryos; *lane 6,* 6–9 hr; *lane 9,* 9–12 hr; *lane nc,* RNA from mass-isolated 9–12 hr nerve cords; *lane 12,* 12–15 hr; *lane 15,* 15–18 hr; *lane 18,* 18–21 hr. The positions of full-length probe (*flp*) and of the products generated by hybridization to RNA of the indicated types are shown to the *left* of the gel.

showed that there are three mRNA types. One of these (type I) lacks the micro-exon and another (type II) contains it. In addition, we found another mRNA class (type III) that contains a second 9 nt insertion immediately 3′ to the SFK micro-exon. This second insertion encodes the amino acids FMN, and we also find this sequence as a separate micro-exon (exon 9) in the genomic fasciclin I sequence. Thus, the three types of mRNA generate proteins that differ by the addition of three and six amino acids to the shortest fasciclin I isoform. We examined the temporal expression pattern of the three mRNA types during

embryonic development, and found that type I mRNA predominates at all times, while type III is present at very low levels. Type II mRNA is expressed primarily between 9 and 15 hr of development and is not found in larvae or adults.

We also examined the structure of grasshopper fasciclin I mRNAs in the D2R region. We found that there are three types of mRNA. One of these corresponds to *Drosophila* type I, and the second has a 9 nt insertion encoding the SFK sequence found in type II mRNA. The third type of grasshopper mRNA (type IV) does not contain the SFK insertion but has instead a 6 nt insertion encoding the amino acids GF. The grasshopper fasciclin I cDNA clone that was originally sequenced is of this type (Zinn et al., 1988). Grasshopper fasciclin I proteins thus differ from the shortest isoform by the addition of three or two additional amino acids. We found no evidence of an mRNA corresponding to *Drosophila* type III, which would contain both insertions. It thus appears that the splicing pattern is different in the two insects and that the FMN micro-exon in *Drosophila* evolved since the two species diverged. Because we have not isolated and sequenced the grasshopper genomic DNA from this region, we do not know how to explain the differences in splicing. The grasshopper gene may have the same organization as in *Drosophila,* with the SFK micro-exon located 5′ to the GF micro-exon. It is also possible that the GF micro-exon is located upstream of the SFK micro-exon. In any event, there is clearly specificity in splicing in both insects, because the *Drosophila* exon 7 splice donor is not used together with the FMN micro-exon splice acceptor, and the GF sequence is not incuded together with the SFK micro-exon in grasshopper mRNAs.

Alternative splicing in cell adhesion molecules

There are numerous examples of alternative splicing of mRNAs encoding cell adhesion molecules. In most of these cases, alternative splicing generates variants with alterations in cytoplasmic domains or transmembrane regions. The best-characterized examples of the generation of extracellular domain variants by alternative splicing are in *Drosophila* PS2α integrin and in vertebrate N-CAM. In PS2α integrin, a 25 amino acid insert is added by alternative splicing between cation binding domains III and IV. This exon is found in PS2α, but not PS1α, and is conserved in PS2α of another species, the Mediterranean fruit fly *Ceratitis capitata* (Brown et al., 1989).

There are at least four isoforms of the extracellular domain of N-CAM, containing insertions at three different sites. The first of these is a muscle-specific form with a 37 amino acid insert in the membrane-proximal region (Dickson et al., 1987; Thompson et al., 1989). A 10 amino acid sequence can be inserted in the fourth immunoglobulin-like domain (Small et al., 1988). Finally, there are inserts of a single and of nine amino acids at the same site in the membrane-proximal region (Santoni et al., 1989).

What is the function of the amino acids encoded by the fasciclin I micro-exons? The D2R sequence is the most conserved segment of the entire fasciclin I protein, suggesting that it serves an important role. The C terminus of D2R, where the micro-exon–encoded amino acids are inserted, forms the junction between the two homologous halves of the protein (domains 1+2, and domains 3+4). McLachlan (1980) has argued that active or binding sites in internally homologous proteins are usually located at the interface between two structural units, and these amino acids are likely to lie at or near this interface. The alteration of D2R by the insertion of three or six amino acids in *Drosophila,* or three or two amino acids in grasshopper, might generate fasciclin I proteins with different binding specificities or strengths.

The importance of just a few amino acids in cell adhesion is not unprecedented. For example, the sequence RGD in fibronectin and other substrate adhesion molecules has been shown to mediate adhesive interactions with the integrin surface receptors, and the context of the RGD sequence may determine the strength of binding to different receptors (reviewed by Ruoslahti and Pierschbacher, 1987). More recently, the sequence LRE has been shown to be an important component of the interaction between motor neurons and the substrate molecule s-laminin (Hunter et al., 1989).

Homophilic adhesion was initially demonstrated for the type II protein (Elkins et al., 1990b) and has since been demonstrated for the type I as well (M. Seeger and C. S. Goodman, unpublished observations). Since the type I protein lacks micro-exon sequences, these amino acids are clearly not essential for adhesion. However, there may be subtle differences in homophilic binding between the two types that cannot be monitored in tissue culture aggregation assays, where cell lines expressing the protein at very high levels are used. Alternatively, it is possible that fasciclin I is also involved in heterophilic interactions *in vivo* and that the micro-exon sequences affect these interactions. The existence of three different protein types might provide the developing nervous system with more specificity by allowing growth cones to distinguish axons bearing different fasciclin I isoforms from each other. The different isoforms could also be used for cell interactions in different tissues. Another possibility is that the inserted amino acids affect protein flexibility or susceptibility to proteolytic degradation.

Of the two 9 bp micro-exons in *Drosophila,* the first, encoding SFK, is conserved between grasshopper and *Drosophila,* while the second, encoding FMN, is not. A different sequence, encoding GF, can be inserted in grasshopper fasciclin I mRNAs. Perhaps these different sequences, FMN versus GF, represent further binding specificities or modifications that have evolved since the two insects diverged 300 million years ago. Alternative micro-exon splicing within a region encoding a binding site could be a useful general mechanism for the evolution of new specificities in adhesion molecules. New sequences capable of functioning as exons that arise within an intron in such a region would occasionally be incorporated into mRNAs. If the resulting protein had a novel activity that was beneficial to the organism, the new sequence might be maintained as an alternatively spliced exon. We hope to define the roles of micro-exon sequences by examining the properties of fasciclin I isoforms in tissue culture experiments, and by transforming *fasI* flies with *fasciclin I* genes bearing mutations in these sequences.

References

Bastiani MJ, Harrelson AL, Snow PM, Goodman CS (1987) Expression of fasciclin I and fasciclin II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. Cell 48: 745–755.

Bender W, Akam M, Karch PA, Peifer M, Spierer P, Lewis EB, Hogness DS (1983) Molecular genetics of the *bithorax*—complex in *Drosophila melanogaster.* Science 221:23–29.

Bieber AJ, Snow PM, Hortsch M, Patel NH, Jacobs JR, Traquina ZR, Schilling J, Goodman CS (1989) *Drosophila* neuroglian: a member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. Cell 59:447–460.

Biggin MD, Tjian R (1989) Transcription factors and the control of *Drosophila* development. Trends Genet 5:377–383.

Bowtell DDL, Simon MA, Rubin GM (1988) Nucleotide sequence and structure of the *sevenless* gene of *Drosophila melanogaster.* Genes & Dev 2:620–634.

Bray SJ, Hirsh J (1986) The *Drosophila virilis* dopa decarboxylase gene is regulated when integrated into *Drosophila melanogaster*. EMBO J 5:2305–2311.

Brown NH, King DL, Wilcox M, Kafatos FC (1989) Developmentally regulated alternative splicing of *Drosophila* integrin PS2-α transcripts. Cell 59:185–195.

Burtis KC, Baker BS (1989) *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding different polypeptides. Cell 56:997–1010.

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995.

Crews ST, Thomas JB, Goodman CS (1988) The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. Cell 52:143–151.

Dickson G, Gower HJ, Barton CH, Prentice HM, Elsom VL, Moore SE, Cox RD, Quinn C, Putt W, Walsh FS (1987) Human muscle neural cell adhesion molecule (NCAM): identification of a muscle-specific sequence in the extracellular domain. Cell 50:1119–1130.

Elkins T, Zinn K, McAllister L, Hoffmann FM, Goodman CS (1990a) Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. Cell 60:565–575.

Elkins T, Hortsch M, Bieber AJ, Snow PM, Goodman CS (1990b) *Drosophila* fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. J Cell Biol 110:1825–1832.

Goodman CS, Bastiani MJ, Doe CQ, du Lac S, Helfand SL, Kuwada JY, Thomas JB (1984) Cell recognition during neuronal development. Science 255:1271–1279.

Grenningloh G, Bieber AJ, Rehm EJ, Snow PM, Traquina Z, Hortsch M, Patel NH, Goodman CS (1990) Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of Ig superfamily cell adhesion molecules. Cold Spring Harbor Symp Quant Biol 55: 327–340.

Harrelson AL, Goodman CS (1988) Growth cone guidance in insects: fasciclin II is a member of the immunoglobulin superfamily. Science 242:700–708.

Hortsch M, Goodman CS (1990) *Drosophila* fasciclin I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. J Biol Chem 265:15104–15109.

Hultmark D, Klemenz R, Gehring W (1986) Translational and transcriptional control elements in the untranslated leader of the heat shock gene *hsp*^{22}. Cell 44:429–438.

Hunter DD, Porter BE, Bulock JW, Adams SP, Merlie JP, Sanes JR (1989) Primary sequence of a motor neuron-selective adhesive site in the synaptic basal lamina protein s-laminin. Cell 59:905–913.

Klambt C, Jacobs JR, Goodman CS (1991) The midline of the *Drosophila* central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance. Cell 64:801–816.

Kruys V, Marinx O, Shaw G, Deschamps J, Huez G (1989) Translation blockade imposed by cytokine-derived UA-rich sequences. Science 245:852–854.

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

McLachlan AD (1980) Pseudo-symmetric structural elements and the folding of domains. In: Protein folding (Jaenicke R, ed), pp. 79–99. Amsterdam: Elsevier/North Holland Biomedical.

Patel NH, Snow PM, Goodman CS (1987) Characterization and cloning of fasciclin III: a glycoprotein expressed on a subset of neurons and axon pathways in *Drosophila*. Cell 48:975–988.

Ruoslahti E, Pierschbacher MD (1987) Arg-Gly-Asp: a versatile cell recognition signal. Cell 44:517–518.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487.

Sanchez-Herrero E, Vernos I, Marco R, Morata G (1985) Genetic organization of *Drosophila* bithorax complex. Nature 313:108–113.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467.

Santoni MJ, Barthels D, Vopper G, Boned A, Goridis C, Wille W (1989) Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. EMBO J 8:385–392.

Shaw G, Kamen R (1986) A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective degradation. Cell 46:659–667.

Small SJ, Haines SL, Akeson RA (1988) Polypeptide variation in an NCAM extracellular immunoglobulin-like fold is developmentally regulated through alternative splicing. Neuron 1:1007–1017.

Snow PM, Zinn K, Harrelson AL, McAllister L, Schilling J, Bastiani MJ, Makk G, Goodman CS (1988) Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. Proc Natl Acad Sci USA 85:5291–5295.

Snow PM, Bieber AJ, Goodman CS (1989) Fasciclin III: a novel homophilic adhesion molecule in *Drosophila*. Cell 59:313–323.

Thompson J, Dickson G, Moore SE, Gower HJ, Putt W, Kenimer JG, Barton CH, Walsh FS (1989) Alternative splicing of the neural cell adhesion molecule gene generates variant extracellular domain structure in skeletal muscle and brain. Genes & Dev 3:348–357.

Zinn K, DiMaio D, Maniatis T (1983) Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell 34: 865–879.

Zinn K, McAllister L, Goodman CS (1988) Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. Cell 53:577–587.